

A Semi-Markovian Modeling of Limit Order Markets

Anatoliy Swishchuk and Nelson Vadori

Department of Mathematics and Statistics, University of Calgary,
2500 University Drive NW, Calgary, Alberta, Canada T2N 1N4
aswish@ucalgary.ca, nvadori@ucalgary.ca

March 11 2015

Abstract. R. Cont and A. de Larrard [5] introduced a tractable stochastic model for the dynamics of a limit order book, computing various quantities of interest such as the probability of a price increase or the diffusion limit of the price process. As suggested by empirical observations, we extend their framework to 1) arbitrary distributions for book events inter-arrival times (possibly non-exponential) and 2) both the nature of a new book event and its corresponding inter-arrival time depend on the nature of the previous book event. We do so by resorting to Markov renewal processes to model the dynamics of the bid and ask queues. We keep analytical tractability via explicit expressions for the Laplace transforms of various quantities of interest. We justify and illustrate our approach by calibrating our model to the five stocks Amazon, Apple, Google, Intel and Microsoft on June 21st 2012. As in [5], the bid-ask spread remains constant equal to one tick, only the bid and ask queues are modeled (they are independent from each other and get reinitialized after a price change), and all orders have the same size.

Key words. limit order book, Markov renewal process, diffusion limit, duration analysis, Weibull, Gamma.

AMS subject classifications. 60K15, 60K20, 90B22, 91B24, 91B70

1 Introduction

Recently, interest in the modeling of limit order markets has increased. Some research has focused on optimal trading strategies in high-frequency environments: for example [8] studies such optimal trading strategies in a context where the stock price follows a semi-Markov process, while market orders arrive in the limit order book via a point process correlated with the stock price itself. [2] develops an optimal execution strategy for an investor seeking to execute a large order using both limit and market orders, under constraints on the volume of such orders. [13] studies optimal execution strategies for the purchase of a large number of shares of a financial asset over a fixed interval of time.

On the other hand, another class of articles has aimed at modeling either the high-frequency dynamics of the stock price itself, or the various queues of outstanding limit orders appearing on the

bid and the ask side of the limit order book, resulting in specific dynamics for the stock price. In [7], a semi-Markov model for the stock price is introduced: the price increments are correlated and equal to arbitrary multiples of the tick size. The correlation between these price increments occurs via their sign only, and not their (absolute) size. In [4], the whole limit order book is modeled (not only the ask and bid queues) via an integer-valued continuous-time Markov chain. Using a Laplace analysis, they compute various quantities of interest such as the probability that the mid-price increases, or the probability that an order is executed before the mid-price moves. A detailed section on parameter estimation is also presented. For a more thorough literature on limit order markets, we refer to the above cited articles and the references thereof.

The starting point of the present manuscript is the article [5], in which a stochastic model for the dynamics of the limit order book is presented. Only the bid and ask queues are modeled (they are independent from each other and get reinitialized after a price change), the bid-ask spread remains constant equal to one tick and all orders have the same size. Their model is analytically tractable, and allows them to compute various quantities of interest such as the distribution of the duration between price changes, the distribution and autocorrelation of price changes, the probability of an upward move in the price and the diffusion limit of the price process. Among the various assumptions made in this article, we seek to challenge two of them while preserving analytical tractability:

- i) the inter-arrival times between book events (limit orders, market orders, order cancellations) are assumed to be independent and exponentially distributed.
- ii) the arrival of a new book event at the bid or the ask is independent from the previous events.

Assumption i) is relatively common among the existing literature ([9], [12], [3], [6], [11], [15], [4]). Nevertheless, as it will be shown later, when calibrating the empirical distributions of the inter-arrival times to the Weibull and Gamma distributions (Amazon, Apple, Google, Intel and Microsoft on June 21st 2012), we find that the shape parameter is in all cases significantly different than 1 (~ 0.1 to 0.3), which suggests that the exponential distribution is typically not rich enough to capture the behavior of these inter-arrival times.

Regarding Assumption ii), we split the book events into 2 different types: limit orders that increase the size of the corresponding bid or ask queue, and market orders/order cancellations that decrease the the size of the corresponding queue. Assimilating the former to the type "+1" and the latter to the type "-1", we find empirically that the probability to get an event of type " ± 1 " is not independent of the nature of the previous event. Indeed, we present below the estimated transition probabilities between book events at the ask and the bid for the stock Microsoft on June 21st 2012. It is seen that the *unconditional* probabilities $P(1)$ and $P(-1)$ to obtain respectively an event of type "+1" and "-1" are relatively close to $1/2$, as in [5]. Nevertheless, denoting $P(i, j)$ the *conditional* probability to obtain an event of type j given that the last event was of type i , we observe that $P(i, j)$ can significantly depend on the previous event i . For example, on the bid side, $P(1, 1) = 0.63$ whereas $P(-1, 1) = 0.36$.

| Microsoft | Bid | Ask |
|------------------|------|------|
| $P(1, 1)$ | 0.63 | 0.60 |
| $P(-1, 1)$ | 0.36 | 0.41 |
| $P(-1, -1)$ | 0.64 | 0.59 |
| $P(1, -1)$ | 0.37 | 0.40 |
| $P(1)$ | 0.49 | 0.51 |
| $P(-1)$ | 0.51 | 0.49 |

Estimated probabilities for book event arrivals. June 21st 2012.

On another front, we will show that we can obtain diffusion limit results for the stock price without resorting to the strong symmetry assumptions of [5]. In particular, the assumption that price increments are i.i.d., which is contrary to empirical observations, as shown in [7] for example.

The paper is organized as follows: section 2 introduces our semi-Markovian modeling of the limit order book, section 3 presents the main probabilistic results obtained in the context of this semi-Markovian model (duration until the next price change, probability of price increase and characterization of the Markov renewal process driving the stock price process), section 4 deals with diffusion limit results for the stock price process, and section 5 presents some calibration results on real market data.

2 A Semi-Markovian modeling of limit order markets

Throughout this paper and to make the reading more convenient, we will use - when appropriate - the same notations as [5], as it is the starting point of the present article. In this section we introduce our model, highlighting when necessary the main differences with the model in [5].

Let s_t , s_t^a , s_t^b be respectively the mid, the ask and the bid price processes. Denoting δ the "tick size", these quantities are assumed to be linked by the following relations:

$$s_t = \frac{1}{2}(s_t^a + s_t^b), \quad s_t^a = s_t^b + \delta.$$

We will also assume that s_0^b is deterministic and positive. In this context, $s_0^a = s_0^b + \delta$ and $s_0 = s_0^b + \frac{\delta}{2}$ are also deterministic and positive. As shown in [5], the assumption that the bid-ask spread $s_t^a - s_t^b$ is constant and equal to one tick does not exactly match the empirical observations, but it is a reasonable assumption as [5] shows that - based on an analysis of the stocks Citigroup, General Electric, General Motors on June 26st 2008 - more than 98% of the observations have a bid-ask spread equal to 1 tick. This corresponds to a situation where the order book contains no empty levels (also called "gaps").

The price process s_t is assumed to be piecewise constant: at random times $\{T_n\}_{n \geq 0}$ (we set $T_0 := 0$), it jumps from its previous value $s_{T_n^-}$ to a new value $s_{T_n} = s_{T_n^-} \pm \delta$. By construction, the same holds for the ask and bid price processes s_t^a and s_t^b . These random times $\{T_n\}$ correspond to the times at which either the bid or the ask queue get depleted, and therefore, the distribution of these

times $\{T_n\}$ will be obtained as a consequence of the dynamics that we will choose to model the bid and ask queues. Let us denote q_t^a and q_t^b the non negative integer-valued processes representing the respective sizes of the ask and bid queues at time t , namely the number of outstanding limit orders at each one of these queues. If the ask queue gets depleted before the bid queue at time T_n - i.e. $q_{T_n}^a = 0$ and $q_{T_n}^b > 0$ - then the price goes up: $s_{T_n} = s_{T_n^-} + \delta$ and both queue values $(q_{T_n}^b, q_{T_n}^a)$ are immediately reinitialized to a new value drawn according to the distribution f , independently from all other random variables. In this context, if n_b, n_a are positive integers, $f(n_b, n_a)$ represents the probability that, after a price increase, the new values of the bid and ask queues are respectively equal to n_b and n_a . On the other hand, if the bid queue gets depleted before the ask queue at time T_n - i.e. $q_{T_n}^a > 0$ and $q_{T_n}^b = 0$ - then the price goes down: $s_{T_n} = s_{T_n^-} - \delta$ and both queue values $(q_{T_n}^b, q_{T_n}^a)$ are immediately reinitialized to a new value drawn according to the distribution \tilde{f} , independently from all other random variables. Following the previous discussion, one can remark that the processes q_t^b, q_t^a will never effectively take the value 0, because whenever $q_{T_n}^b = 0$ or $q_{T_n}^a = 0$, we "replace" the pair $(q_{T_n}^b, q_{T_n}^a)$ by a random variable drawn from the distribution f or \tilde{f} . The precise construction of the processes (q_t^b, q_t^a) will be explained below.

Let $\tau_n := T_n - T_{n-1}$ the "sojourn times" between two consecutive price changes, $N_t := \sup\{n : T_n \leq t\} = \sup\{n : \tau_1 + \dots + \tau_n \leq t\}$ the number of price changes up to time t , $X_n := s_{T_n} - s_{T_{n-1}}$ the consecutive price increments (which can only take the values $\pm\delta$). With these notations we have:

$$s_t = \sum_{k=1}^{N_t} X_k.$$

Let us now present the chosen model for the dynamics of the bid and ask queues. As mentioned in introduction, we seek to extend the model [5] in the two following directions, as suggested by our calibration results:

- i) inter-arrival times between book events (limit orders, market orders, order cancellations) are allowed to have an arbitrary distribution.
- ii) the arrival of a new book event at the bid or the ask and its corresponding inter-arrival time are allowed to depend on the nature of the previous event.

In order to do so, we will use a Markov renewal structure for the joint process of book events and corresponding inter-arrival times occurring at the ask and bid sides. Formally, for the ask side, consider a family $\{R^{n,a}\}_{n \geq 0}$ of Markov renewal processes given by:

$$R^{n,a} := \{(V_k^{n,a}, T_k^{n,a})\}_{k \geq 0}.$$

For each n , $R^{n,a}$ will "drive" the dynamics of the ask queue on the interval $[T_n, T_{n+1})$ where the stock price remains constant. $\{V_k^{n,a}\}_{k \geq 0}$ and $\{T_k^{n,a}\}_{k \geq 0}$ represent respectively the consecutive book events and the consecutive inter-arrival times between these book events at the ask side on the interval $[T_n, T_{n+1})$. At time T_{n+1} where one of the bid or ask queues gets depleted, the stock price changes and the model will be reinitialized with an independent copy $R^{n+1,a}$ of $R^{n,a}$: it will therefore be assumed that the processes $\{R^{n,a}\}_{n \geq 0}$ are independent copies of the same Markov

renewal process of kernel Q^a , namely for each n :

$$\begin{aligned}\mathbb{P}[V_{k+1}^{n,a} = j, T_{k+1}^{n,a} \leq t | T_p^{n,a}, V_p^{n,a} : p \leq k] &= Q^a(V_k^{n,a}, j, t), \quad j \in \{-1, 1\} \\ \mathbb{P}[V_0^{n,a} = j] &= v_0^a(j), \quad j \in \{-1, 1\} \\ \mathbb{P}[T_0^{n,a} = 0] &= 1.\end{aligned}$$

We recall that as mentioned earlier, we consider two types of book events $V_k^{n,a}$: events of type $+1$ which increase the ask queue by 1 (limit orders), and events of type -1 which decrease the ask queue by 1 (market orders and order cancellations). In particular, the latter assumptions constitute a generalization of [5] in the sense that for each n :

- $V_{k+1}^{n,a}$ depends on the previous queue change $V_k^{n,a}$: $\{V_k^{n,a}\}_{k \geq 0}$ is a Markov chain.
- the inter-arrival times $\{T_k^{n,a}\}_{k \geq 0}$ between book events can have arbitrary distributions. Further, they are not strictly independent anymore but they are independent conditionally on the Markov chain $\{V_k^{n,a}\}_{k \geq 0}$.

We use the same notations to model the bid queue - but with indexes a replaced by b - and we assume that the processes involved at the bid and at the ask are independent.

In [5], the kernel Q^a is given by (the kernel Q^b has a similar expression with indexes a replaced by b):

$$\begin{aligned}Q^a(i, 1, t) &= \frac{\lambda^a}{\lambda^a + \theta^a + \mu^a} (1 - e^{-(\lambda^a + \theta^a + \mu^a)t}), \quad i \in \{-1, 1\} \\ Q^a(i, -1, t) &= \frac{\theta^a + \mu^a}{\lambda^a + \theta^a + \mu^a} (1 - e^{-(\lambda^a + \theta^a + \mu^a)t}), \quad i \in \{-1, 1\}.\end{aligned}$$

Given these chosen dynamics to model to ask and bid queues between two consecutive price changes, we now specify formally the "state process":

$$\tilde{L}_t := (s_t^b, q_t^b, q_t^a)$$

which will keep track of the state of the limit order book at time t (stock price and sizes of the bid and ask queues). In the context of [5], this process \tilde{L}_t was proved to be Markovian. Here, we will need to "add" to this process the process (V_t^b, V_t^a) keeping track of the nature of the last book event at the bid and the ask to make it Markovian: in this sense we can view it as being semi-Markovian. The process:

$$L_t := (s_t^b, q_t^b, q_t^a, V_t^b, V_t^a)$$

constructed below will be proved to be Markovian.

The process L is piecewise constant and changes value whenever a book event occurs at the bid or at the ask. We will construct both the process L and the sequence of times $\{T_n\}_{n \geq 0}$ recursively on $n \geq 0$. The recursive construction starts from $n = 0$ where we have $T_0 = 0$, $s_0^b > 0$ deterministic, and $(q_0^b, q_0^a, V_0^b, V_0^a)$ is a random variable with distribution $f_0 \times v_0^b \times v_0^a$, where f_0 is a distribution on $\mathbb{N}^* \times \mathbb{N}^*$, and both v_0^b and v_0^a are distributions on the two-point space $\{-1, 1\}$, that is $v_0^b(1) =$

$\mathbb{P}[V_0^b = 1]$ is given and $v_0^b(-1) = 1 - v_0^b(1)$ (and similarly for the ask). We will need to introduce the following processes for the bid side (for the ask side, they are defined similarly):

$$\bar{T}_k^{n,b} := \sum_{p=0}^k T_p^{n,b}, \quad N_t^{n,b} := \sup\{k : T_n + \bar{T}_k^{n,b} \leq t\}.$$

With these notations, the book events corresponding to the interval $[T_n, T_{n+1})$ occur at times $T_n + \bar{T}_k^{n,b}$ ($k \geq 0$) until one of the queues gets depleted, and $N_t^{n,b}$ counts the number of book events on the interval $[T_n, t]$, for $t \in [T_n, T_{n+1})$.

The joint construction of L and of the sequence of times $\{T_n\}_{n \geq 0}$ is done recursively on $n \geq 0$. The following describes the step n of the recursive construction:

- For each $T \in \{T_n + \bar{T}_k^{n,b}\}_{k \geq 1}$, the book event $v_{n,T}^b := V_{N_T^{n,b}}^{n,b}$ occurs at time T at the bid side. If $q_{T-}^b + v_{n,T}^b > 0$, there is no price change at time T and we have:

$$(s_T^b, q_T^b, q_T^a, V_T^b, V_T^a) = (s_{T-}^b, q_{T-}^b + v_{n,T}^b, q_{T-}^a, v_{n,T}^b, V_{T-}^a).$$

If on the other hand $q_{T-}^b + v_{n,T}^b = 0$, there is a price change at time T and the model gets reinitialized:

$$(s_T^b, q_T^b, q_T^a, V_T^b, V_T^a) = (s_{T-}^b - \delta, \tilde{x}_n^b, \tilde{x}_n^a, v_{0,n}^b, v_{0,n}^a),$$

where $\{(\tilde{x}_k^b, \tilde{x}_k^a)\}_{k \geq 0}$ are i.i.d. random variables, independent from all other random variables, with joint distribution \tilde{f} on $\mathbb{N}^* \times \mathbb{N}^*$, and $\{v_{0,k}^b, v_{0,k}^a\}_{k \geq 0}$ are i.i.d. random variables, independent from all other random variables, with joint distribution $v_0^b \times v_0^a$ on the space $\{-1, 1\} \times \{-1, 1\}$. We then set $T_{n+1} = T$ and move from the step n of the recursion to the step $n + 1$.

- For each $T \in \{T_n + \bar{T}_k^{n,a}\}_{k \geq 1}$, the book event $v_{n,T}^a := V_{N_T^{n,a}}^{n,a}$ occurs at time T at the ask side. If $q_{T-}^a + v_{n,T}^a > 0$, there is no price change at time T and we have:

$$(s_T^b, q_T^b, q_T^a, V_T^b, V_T^a) = (s_{T-}^b, q_{T-}^b, q_{T-}^a + v_{n,T}^a, V_{T-}^b, v_{n,T}^a).$$

If on the other hand $q_{T-}^a + v_{n,T}^a = 0$, there is a price change at time T and the model gets reinitialized:

$$(s_T^b, q_T^b, q_T^a, V_T^b, V_T^a) = (s_{T-}^b + \delta, x_n^b, x_n^a, v_{0,n}^b, v_{0,n}^a),$$

where $\{(x_k^b, x_k^a)\}_{k \geq 0}$ are i.i.d. random variables, independent from all other random variables, with joint distribution f on $\mathbb{N}^* \times \mathbb{N}^*$, and $\{v_{0,k}^b, v_{0,k}^a\}_{k \geq 0}$ are the i.i.d. random variables defined above. We then set $T_{n+1} = T$ and move from the step n of the recursion to the step $n + 1$.

It results from the above construction and the Markov renewal structure of the processes $\{R^{n,a}\}_{n \geq 0}$, $\{R^{n,b}\}_{n \geq 0}$ that the process L_t is Markovian.

Since the processes $\{R^{n,a}\}_{n \geq 0}$ are independent copies of the same Markov renewal process of kernel Q^a , we will drop the index n when appropriate in order to make the notations lighter. Following this remark, we will introduce the following notations for the ask, for $i, j \in \{-1, 1\}$ (for the bid, they are defined similarly):

$$\begin{aligned}
P^a(i, j) &:= \mathbb{P}[V_{k+1}^a = j | V_k^a = i], \\
F^a(i, t) &:= \mathbb{P}[T_{k+1}^a \leq t | V_k^a = i], \\
H^a(i, j, t) &:= \mathbb{P}[T_{k+1}^a \leq t | V_k^a = i, V_{k+1}^a = j], \\
h^a(i, j) &:= \int_0^\infty t H^a(i, j, dt), \\
h_1^a &:= h^a(1, 1) + h^a(-1, -1), \quad h_2^a := h^a(-1, 1) + h^a(1, -1), \\
m^a(s, i, j) &:= \int_0^\infty e^{-st} Q^a(i, j, dt), \quad s \in \mathbb{C}, \\
M^a(s, i) &:= m^a(s, i, -1) + m^a(s, i, 1) = \int_0^\infty e^{-st} F^a(i, dt), \quad s \in \mathbb{C}.
\end{aligned}$$

Throughout this paper, we will use the following mild technical assumptions:

$$\textbf{(A1)} \quad 0 < P^a(i, j) < 1, \quad 0 < P^b(i, j) < 1, \quad i, j \in \{-1, 1\}.$$

$$\textbf{(A2)} \quad F^a(i, 0) < 1, \quad F^b(i, 0) < 1, \quad i \in \{-1, 1\}.$$

$$\textbf{(A3)} \quad \int_0^\infty t^2 H^a(i, j, dt) < \infty, \quad \int_0^\infty t^2 H^b(i, j, dt) < \infty, \quad i, j \in \{-1, 1\}.$$

Some brief comments on these assumptions: **(A1)** implies that each state ± 1 is accessible from each state. **(A2)** means that each inter-arrival time between book events has a positive probability to be non zero, and **(A3)** constitutes a second moment integrability assumption on the cumulative distribution functions H^a and H^b .

3 Main Probabilistic Results

Throughout this section and as mentioned earlier, since the processes $\{R^{n,a}\}_{n \geq 0}$ are independent copies of the same Markov renewal process of kernel Q^a , we will drop the index n when appropriate in order to make the notations lighter on the random variables $T_k^{n,a}$, $\bar{T}_k^{n,a}$, $V_k^{n,a}$ (and similarly for the bid side).

3.1 Duration until the next price change

Given an initial configuration of the bid and ask queues $(q_0^b, q_0^a) = (n_b, n_a)$ (n_b, n_a integers), we denote σ_b the first time at which the bid queue is depleted:

$$\sigma_b = \bar{T}_{k^*}^b, \quad k^* := \inf \left\{ k : n_b + \sum_{m=1}^k V_m^b = 0 \right\}.$$

Similarly we define σ_a the first time at which the ask queue is depleted. The duration until the next price move is thus:

$$\tau := \sigma_a \wedge \sigma_b.$$

In order to have a realistic model in which the queues always get depleted at some point, i.e. $\mathbb{P}[\sigma_a < \infty] = \mathbb{P}[\sigma_b < \infty] = 1$, we impose the conditions:

$$P^a(1, 1) \leq P^a(-1, -1), \quad P^b(1, 1) \leq P^b(-1, -1).$$

These conditions correspond to the condition $\lambda \leq \theta + \mu$ in [5], and the proof of the proposition below shows that they are respectively equivalent to $\mathbb{P}[\sigma_a < \infty] = 1$ and $\mathbb{P}[\sigma_b < \infty] = 1$. Indeed, as $s \rightarrow 0$ ($s > 0$), the Laplace transform $\mathcal{L}^a(s) := \mathbb{E}[e^{-s\sigma_a}]$ of σ_a tends to $\mathbb{P}[\sigma_a < \infty]$. The proposition below shows that if $P^a(1, 1) > P^a(-1, -1)$, this quantity is strictly less than 1, and if $P^a(1, 1) \leq P^a(-1, -1)$, this quantity is equal to 1. We have the following result which generalizes the Proposition 1 in [5] (see also Remark .2 below):

Proposition .1. *The conditional law of σ_a given $q_0^a = n \geq 1$ has a regularly varying tail with:*

- *tail exponent 1 if $P^a(1, 1) < P^a(-1, -1)$.*
- *tail exponent 1/2 if $P^a(1, 1) = P^a(-1, -1)$.*

More precisely, we get: if $P^a(1, 1) = P^a(-1, -1) = p_a$:

$$\mathbb{P}[\sigma_a > t | q_0^a = n] \stackrel{t \rightarrow \infty}{\sim} \frac{\alpha^a(n)}{\sqrt{t}}$$

with:

$$\alpha^a(n) := \frac{1}{p_a \sqrt{\pi}} \left(n + \frac{2p_a - 1}{p_a - 1} v_0^a(1) \right) \sqrt{p_a(1 - p_a)} \sqrt{p_a h_1^a + (1 - p_a) h_2^a}.$$

If $P^a(1, 1) < P^a(-1, -1)$, we get:

$$\mathbb{P}[\sigma_a > t | q_0^a = n] \stackrel{t \rightarrow \infty}{\sim} \frac{\beta^a(n)}{t}$$

with:

$$\begin{aligned} \beta^a(n) &:= v_0^a(1) u_1^a + v_0^a(-1) u_2^a + (n - 1) u_3^a, \\ u_1^a &:= h^a(1, -1) + \frac{P^a(1, 1)}{1 - P^a(1, 1)} (u_3^a + h^a(1, 1)) \\ u_2^a &:= -h^a(1, 1) + \frac{1 - P^a(-1, -1)}{1 - P^a(1, 1)} (u_3^a + h^a(1, 1)) + P^a(-1, -1) h_1^a + (1 - P^a(-1, -1)) h_2^a, \\ u_3^a &:= h^a(1, 1) + \frac{1 - P^a(1, 1)}{P^a(-1, -1) - P^a(1, 1)} (P^a(-1, -1) h_1^a + (1 - P^a(-1, -1)) h_2^a). \end{aligned}$$

Similar expressions are obtained for $\mathbb{P}[\sigma_b > t | q_0^b = n]$, with indexes a replaced by b .

Remark .2. We retrieve the results of [5]: if $P^a(1, 1) = P^a(-1, -1)$, then within the context/notations of [5] we get $p_a = 1/2$ and:

$$h^a(i, j) = \int_0^\infty 2t\lambda e^{-2\lambda t} dt = \frac{1}{2\lambda},$$

and so $\alpha^a(n) = \frac{n}{\sqrt{\pi\lambda}}$. For the case $P^a(1, 1) < P^a(-1, -1)$ ($\lambda < \theta + \mu$ with their notations), we find:

$$\beta^a(n) = \frac{n}{\theta + \mu - \lambda},$$

which is different from the result of [5] that is $\beta^a(n) = \frac{n(\theta + \mu + \lambda)}{2\lambda(\theta + \mu - \lambda)}$. We believe that they made a small mistake in their Taylor expansion on page 10: in the case $\lambda < \theta + \mu$, they should find:

$$\mathcal{L}(s, x) \xrightarrow{s \rightarrow 0} 1 - \frac{sx}{\theta + \mu - \lambda}.$$

Proof. Let $s > 0$ and denote $\mathcal{L}(s, n, i) := \mathbb{E}[e^{-s\sigma_a} | q_0^a = n, V_0^a = i]$. We have:

$$\sigma_a = \sum_{m=1}^{k^*} T_m^a, \quad k^* := \inf\{k : n + \sum_{m=1}^k V_m^a = 0\}.$$

Therefore:

$$\begin{aligned} \mathcal{L}(s, n, i) &= \mathbb{E}[e^{-sT_1^a} \mathbb{E}[e^{-s(\sigma_a - T_1^a)} | q_0^a = n, V_0^a = i, V_1^a, T_1^a] | q_0^a = n, V_0^a = i] \\ &= \mathbb{E}[e^{-sT_1^a} \underbrace{\mathbb{E}[e^{-s(\sigma_a - T_1^a)} | q_{T_1^a}^a = n + V_1^a, V_0^a = i, V_1^a, T_1^a]}_{\mathcal{L}(s, n + V_1^a, V_1^a)} | q_0^a = n, V_0^a = i] \\ &= \mathbb{E}[e^{-sT_1^a} \mathcal{L}(s, n + V_1^a, V_1^a) | q_0^a = n, V_0^a = i] \\ &= \int_0^\infty e^{-st} \mathcal{L}(s, n + 1, 1) Q^a(i, 1, dt) + \int_0^\infty e^{-st} \mathcal{L}(s, n - 1, -1) Q^a(i, -1, dt) \\ &= m^a(s, i, 1) \mathcal{L}(s, n + 1, 1) + m^a(s, i, -1) \mathcal{L}(s, n - 1, -1) \end{aligned}$$

Denote for sake of clarity $a_n := \mathcal{L}(s, n, 1)$, $b_n := \mathcal{L}(s, n, -1)$. These sequences therefore solve the system of coupled recurrence equations:

$$\begin{aligned} a_{n+1} &= m^a(s, 1, 1) a_{n+2} + m^a(s, 1, -1) b_n, \quad n \geq 0 \\ b_{n+1} &= m^a(s, -1, 1) a_{n+2} + m^a(s, -1, -1) b_n \\ a_0 &= b_0 = 1. \end{aligned}$$

Simple algebra (computing $a_{n+1} - m^a(s, -1, -1) a_n$ on the one hand and $m^a(s, 1, 1) b_{n+1} - b_n$ on the other hand) gives us that both a_n and b_n solve the same following recurrence equation (but for different initial conditions):

$$m^a(s, 1, 1) u_{n+2} - (1 + \Delta^a(s)) u_{n+1} + m^a(s, -1, -1) u_n, \quad n \geq 1$$

with:

$$\Delta^a(s) := m^a(s, 1, 1)m^a(s, -1, -1) - m^a(s, -1, 1)m^a(s, 1, -1).$$

The parameter $\Delta^a(s)$ can be seen as a coupling coefficient and is equal to 0 when the random variable (V_k^a, T_k^a) doesn't depend on the previous state V_{k-1}^a , for example in the context of [5].

If we denote $R(X)$ the characteristic polynomial associated to the previous recurrence equation $R(X) := m^a(s, 1, 1)X^2 - (1 + \Delta^a(s))X + m^a(s, -1, -1)$, then simple algebra gives us:

$$R(1) = \underbrace{(M^a(s, 1) - 1)}_{<0} \underbrace{(1 - m^a(s, -1, -1))}_{>0} + \underbrace{m^a(s, 1, -1)}_{>0} \underbrace{(M^a(s, -1) - 1)}_{<0} < 0$$

Note that $M^a(s, i) < 1$ for $s > 0$ because $F^a(i, 0) < 1$. Since $m^a(s, 1, 1) > 0$, this implies that R has only one root < 1 (and an other root > 1):

$$\lambda^a(s) := \frac{1 + \Delta^a(s) - \sqrt{(1 + \Delta^a(s))^2 - 4m^a(s, 1, 1)m^a(s, -1, -1)}}{2m^a(s, 1, 1)}.$$

Because we have $a_n, b_n \leq 1$ for $s > 0$, then we must have for $n \geq 1$:

$$a_n = a_1 \lambda^a(s)^{n-1} \qquad b_n = b_1 \lambda^a(s)^{n-1}$$

The recurrence equations on a_n, b_n give us:

$$a_1 = \frac{m^a(s, 1, -1)}{1 - \lambda^a(s)m^a(s, 1, 1)} \qquad b_1 = \frac{m^a(s, -1, 1)a_1 + \Delta^a(s)}{m^a(s, 1, 1)}$$

Finally, letting $\mathcal{L}(s, n) := \mathbb{E}[e^{-s\sigma_a} | q_0^a = n]$, we obtain:

$$\mathcal{L}(s, n) = \sum_i \mathcal{L}(s, n, i) v_0^a(i) = a_n v_0^a(1) + b_n v_0^a(-1).$$

The behavior of $\mathbb{P}[\sigma_a > t | q_0^a = n]$ as $t \rightarrow \infty$ is obtained by computing the behavior of $\mathcal{L}(s, n)$ as $s \rightarrow 0$, together with Karamata's Tauberian theorem. By the second moment integrability assumption on $H^a(i, j, dt)$, we note that:

$$\begin{aligned} m^a(s, i, j) &= \int_0^\infty e^{-st} Q^a(i, j, dt) = P^a(i, j) \int_0^\infty e^{-st} H^a(i, j, dt) \\ &\stackrel{s \rightarrow 0}{\sim} P^a(i, j) - s P^a(i, j) \int_0^\infty t H^a(i, j, dt) = P^a(i, j) - s P^a(i, j) h^a(i, j). \end{aligned}$$

Now, assume $P^a(1, 1) = P^a(-1, -1) = p_a$. A straightforward but tedious Taylor expansion of $\mathcal{L}(s, n)$ as $s \rightarrow 0$ gives us:

$$\mathcal{L}(s, n) \stackrel{s \rightarrow 0}{\sim} 1 - \sqrt{\pi} \alpha^a(n) \sqrt{s}.$$

The same way, if $P^a(1, 1) < P^a(-1, -1)$, a straightforward Taylor expansion of $\mathcal{L}(s, n)$ as $s \rightarrow 0$ gives us:

$$\mathcal{L}(s, n) \stackrel{s \rightarrow 0}{\sim} 1 - \beta^a(n)s.$$

□

We are interested in the asymptotic behavior of the law of τ , which is, by independence of the bid/ask queues:

$$\mathbb{P}[\tau > t | (q_0^b, q_0^a) = (n_b, n_a)] = \mathbb{P}[\sigma_a > t | q_0^a = n_a] \mathbb{P}[\sigma_b > t | q_0^b = n_b].$$

We get the following immediate consequence of Proposition .1:

Proposition .3. *The conditional law of τ given $(q_0^b, q_0^a) = (n_b, n_a)$ has a regularly varying tail with:*

- *tail exponent 2 if $P^a(1, 1) < P^a(-1, -1)$ and $P^b(1, 1) < P^b(-1, -1)$. In particular, in this case, $\mathbb{E}[\tau | (q_0^b, q_0^a) = (n_b, n_a)] < \infty$.*
- *tail exponent 1 if $P^a(1, 1) = P^a(-1, -1)$ and $P^b(1, 1) = P^b(-1, -1)$. In particular, in this case, $\mathbb{E}[\tau | (q_0^b, q_0^a) = (n_b, n_a)] = \infty$ whenever $n_b, n_a \geq 1$.*
- *tail exponent 3/2 otherwise. In particular, in this case, $\mathbb{E}[\tau | (q_0^b, q_0^a) = (n_b, n_a)] < \infty$.*

More precisely, we get: if $P^a(1, 1) = P^a(-1, -1)$ and $P^b(1, 1) = P^b(-1, -1)$:

$$\mathbb{P}[\tau > t | (q_0^b, q_0^a) = (n_b, n_a)] \stackrel{t \rightarrow \infty}{\sim} \frac{\alpha^a(n_a) \alpha^b(n_b)}{t}$$

if $P^a(1, 1) < P^a(-1, -1)$ and $P^b(1, 1) < P^b(-1, -1)$:

$$\mathbb{P}[\tau > t | (q_0^b, q_0^a) = (n_b, n_a)] \stackrel{t \rightarrow \infty}{\sim} \frac{\beta^a(n_a) \beta^b(n_b)}{t^2}$$

if $P^a(1, 1) = P^a(-1, -1)$ and $P^b(1, 1) < P^b(-1, -1)$:

$$\mathbb{P}[\tau > t | (q_0^b, q_0^a) = (n_b, n_a)] \stackrel{t \rightarrow \infty}{\sim} \frac{\alpha^a(n_a) \beta^b(n_b)}{t^{3/2}}$$

if $P^a(1, 1) < P^a(-1, -1)$ and $P^b(1, 1) = P^b(-1, -1)$:

$$\mathbb{P}[\tau > t | (q_0^b, q_0^a) = (n_b, n_a)] \stackrel{t \rightarrow \infty}{\sim} \frac{\beta^a(n_a) \alpha^b(n_b)}{t^{3/2}}$$

Proof. Immediate using proposition .1.

□

It will be needed to get the full law of τ , which is, by independence of the bid/ask queues:

$$\mathbb{P}[\tau > t | (q_0^b, q_0^a) = (n_b, n_a)] = \mathbb{P}[\sigma_a > t | q_0^a = n_a] \mathbb{P}[\sigma_b > t | q_0^b = n_b].$$

We have computed explicitly the Laplace transforms of σ_a and σ_b (cf. the proof of Proposition .1 above). There are two possibilities: either it is possible to invert those Laplace transforms so that we can compute $\mathbb{P}[\sigma_a > t | q_0^a = n_a]$ and $\mathbb{P}[\sigma_b > t | q_0^b = n_b]$ in closed form and thus $\mathbb{P}[\tau > t | (q_0^b, q_0^a) = (n_b, n_a)]$ in closed form as in [5]. If not, we will have to resort to a numerical procedure to invert the characteristic functions of σ_a and σ_b . Below we give the characteristic functions of σ_a and σ_b :

Proposition .4. *Let $\phi^a(t, n) := \mathbb{E}[e^{it\sigma_a} | q_0^a = n]$ ($t \in \mathbb{R}$) the characteristic function of σ_a conditionally on $q_0^a = n \geq 1$. We have:*

if $m^a(-it, 1, 1) \neq 0$:

$$\begin{aligned} \phi^a(t, n) &= (c^a(-it)v_0^a(1) + d^a(-it)v_0^a(-1)) \lambda^a(-it)^{n-1}, \\ c^a(z) &= \frac{m^a(z, 1, -1)}{1 - \lambda^a(z)m^a(z, 1, 1)}, \\ d^a(z) &= \frac{m^a(z, -1, 1)c^a(z) + \Delta^a(z)}{m^a(z, 1, 1)}, \\ \Delta^a(z) &:= m^a(z, 1, 1)m^a(z, -1, -1) - m^a(z, -1, 1)m^a(z, 1, -1), \\ \lambda^a(z) &:= \frac{1 + \Delta^a(z) - \sqrt{(1 + \Delta^a(z))^2 - 4m^a(z, 1, 1)m^a(z, -1, -1)}}{2m^a(z, 1, 1)}. \end{aligned}$$

and if $m^a(-it, 1, 1) = 0$:

$$\begin{aligned} \phi^a(t, n) &= (m^a(-it, 1, -1)v_0^a(1) + \widetilde{\lambda}^a(-it)v_0^a(-1))\widetilde{\lambda}^a(-it)^{n-1}, \\ \widetilde{\lambda}^a(z) &:= \frac{m^a(z, -1, -1)}{1 - m^a(z, 1, -1)m^a(z, -1, 1)}. \end{aligned}$$

The coefficient $\Delta^a(z)$ can be seen as a coupling coefficient and is equal to 0 when the random variable (V_k^a, T_k^a) doesn't depend on the previous state V_{k-1}^a , for example in the context of [5].

The characteristic function $\phi^b(t, n) := \mathbb{E}[e^{it\sigma_b} | q_0^b = n]$ has the same expression, with indexes a replaced by b .

Proof. Similarly to the proof of Proposition .1, we obtain (using the same notations but denoting this time $a_n := \mathcal{L}(-it, n, 1)$, $b_n := \mathcal{L}(-it, n, -1)$):

$$\begin{aligned} a_{n+1} &= m^a(-it, 1, 1)a_{n+2} + m^a(-it, 1, -1)b_n, \quad n \geq 0 \\ b_{n+1} &= m^a(-it, -1, 1)a_{n+2} + m^a(-it, -1, -1)b_n \\ a_0 &= b_0 = 1. \end{aligned}$$

If $m^a(-it, 1, 1) = 0$, we can solve explicitly the above system to get the desired result. If $m^a(-it, 1, 1) \neq 0$, we get as in the proof of Prop .1 that both a_n and b_n solve the same following recurrence equation (but for different initial conditions):

$$m^a(-it, 1, 1)u_{n+2} - (1 + \Delta^a(-it))u_{n+1} + m^a(-it, -1, -1)u_n, \quad n \geq 1.$$

Because $|m^a(-it, j, -1) + m^a(-it, j, 1)| = |M^a(-it, j)| = \left| \int_0^\infty e^{its} F^a(j, ds) \right| \leq 1$, tedious computations give us that $|\lambda_+^a(-it)| > 1$ whenever $t \neq 0$, where:

$$\lambda_+^a(z) := \frac{1 + \Delta^a(z) + \sqrt{(1 + \Delta^a(z))^2 - 4m^a(z, 1, 1)m^a(z, -1, -1)}}{2m^a(z, 1, 1)}.$$

Since both $|a_n|, |b_n| \leq 1$ for all n , it must be that:

$$a_n = a_1 \lambda_+^a(-it)^{n-1} \quad b_n = b_1 \lambda_+^a(-it)^{n-1},$$

with a_1, b_1 being given by the recurrence equations on a_n, b_n :

$$a_1 = \frac{m^a(-it, 1, -1)}{1 - \lambda_+^a(-it)m^a(-it, 1, 1)} \quad b_1 = \frac{m^a(-it, -1, 1)a_1 + \Delta^a(-it)}{m^a(-it, 1, 1)}.$$

Finally we conclude by observing that:

$$\phi^a(t, n) = a_n v_0^a(1) + b_n v_0^a(-1).$$

□

3.2 Probability of Price Increase

Starting from an initial configuration of the bid and ask queues, $(q_0^b, q_0^a) = (n_b, n_a)$, the probability that the next price change is a price increase will be denoted $p_1^{up}(n_b, n_a)$. This quantity is equal to the probability that σ_a is less than σ_b :

$$p_1^{up}(n_b, n_a) = \mathbb{P}[\sigma_a < \sigma_b | q_0^b = n_b, q_0^a = n_a].$$

Since we know the characteristic functions of σ_a, σ_b (cf. Proposition .4), we can compute their individual laws up to the use of a numerical procedure. Since σ_a and σ_b are independent, the law of $\sigma_b - \sigma_a$ can be computed using the individual laws of σ_a, σ_b , and therefore $p_1^{up}(n_b, n_a)$ can be computed up to the use of numerical procedures to 1) invert the characteristic function and 2) compute an indefinite integral. Indeed, denoting $f_{n_a, a}$ the p.d.f of σ_a conditionally on $q_0^a = n_a$, and $F_{n_b, b}$ the c.d.f. of σ_b conditionally on $q_0^b = n_b$, we have:

$$p_1^{up}(n_b, n_a) = \mathbb{P}[\sigma_a < \sigma_b | q_0^b = n_b, q_0^a = n_a] = \int_0^\infty f_{n_a, a}(t)(1 - F_{n_b, b}(t))dt,$$

where $F_{n_b,b}$ and $f_{n_a,a}$ are obtained by the following inversion formulas:

$$\begin{aligned} f_{n_a,a}(t) &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \phi^a(x, n_a) dx, \\ F_{n_b,b}(t) &= \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{1}{x} \text{Im}\{e^{-itx} \phi^b(x, n_b)\} dx. \end{aligned}$$

3.3 The stock price seen as a functional of a Markov renewal process

As mentioned earlier, we can write the stock price s_t as:

$$s_t = \sum_{k=1}^{N_t} X_k,$$

where $\{X_n\}_{n \geq 0}$ are the consecutive price increments taking value $\pm\delta$, $\{\tau_n\}_{n \geq 0}$ are the consecutive durations between price changes and $\{T_n\}_{n \geq 0}$ the consecutive times at which the price changes.

In this context, the distribution of the random variable τ_{n+1} will depend on the initial configuration of the bid and ask queues at the beginning T_n of the period $[T_n, T_{n+1})$, which itself depends on the nature of the previous price change X_n : if the previous price change is a price decrease, the initial configuration will be drawn from the distribution \tilde{f} , and if it is an increase, the initial configuration will be drawn from the distribution f . Because for each n the random variable (X_n, τ_n) only depends on the previous increment X_{n-1} , it can be seen that the process $(X_n, \tau_n)_{n \geq 0}$ is a Markov renewal process ([10], [16]), and the stock price can therefore be seen as a functional of this Markov renewal process. We obtain the following result.

Proposition .5. *The process $(X_n, \tau_n)_{n \geq 0}$ is a Markov renewal process. The law of the process $\{\tau_n\}_{n \geq 0}$ is given by:*

$$\begin{aligned} F(\delta, t) &:= \mathbb{P}[\tau_{n+1} \leq t | X_n = \delta] = \sum_{p=1}^{\infty} \sum_{n=1}^{\infty} f(n, p) \mathbb{P}[\tau \leq t | (q_0^b, q_0^a) = (n, p)], \\ F(-\delta, t) &:= \mathbb{P}[\tau_{n+1} \leq t | X_n = -\delta] = \sum_{p=1}^{\infty} \sum_{n=1}^{\infty} \tilde{f}(n, p) \mathbb{P}[\tau \leq t | (q_0^b, q_0^a) = (n, p)]. \end{aligned}$$

The Markov chain $\{X_n\}_{n \geq 0}$ is characterized by the following transition probabilities:

$$\begin{aligned} p_{cont} &:= \mathbb{P}[X_{n+1} = \delta | X_n = \delta] = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p_1^{up}(i, j) f(i, j). \\ p'_{cont} &:= \mathbb{P}[X_{n+1} = -\delta | X_n = -\delta] = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} (1 - p_1^{up}(i, j)) \tilde{f}(i, j). \end{aligned}$$

The generator of this Markov chain is thus (we assimilate the state 1 to the value δ and the state 2 to the value $-\delta$):

$$P := \begin{pmatrix} p_{cont} & 1 - p_{cont} \\ 1 - p'_{cont} & p'_{cont} \end{pmatrix}$$

Let $p_n^{up}(b, a) := \mathbb{P}[X_n = \delta | q_0^b = b, q_0^a = a]$. We can compute this quantity explicitly:

$$p_n^{up}(b, a) = \pi^* + (p_{cont} + p'_{cont} - 1)^{n-1} (p_1^{up}(b, a) - \pi^*),$$

$$\pi^* := \pi^*(\delta) := \frac{p'_{cont} - 1}{p_{cont} + p'_{cont} - 2},$$

where π^* is the stationary distribution of the Markov chain $\{X_n\}$:

$$\pi^* = \lim_{n \rightarrow \infty} \mathbb{P}[X_n = \delta | X_1].$$

Further:

$$\mathbb{E}[X_n | q_0^b = b, q_0^a = a] = \delta(2p_n^{up}(b, a) - 1),$$

and the (conditional) covariance between two consecutive price moves:

$$\text{cov}[X_{n+1}, X_n | q_0^b = b, q_0^a = a] = 4\delta^2 p_n^{up}(b, a)(1 - p_n^{up}(b, a))(p_{cont} + p'_{cont} - 1).$$

Remark .6. In particular, if $p_{cont} = p'_{cont}$, then $\pi^* = 1/2$ and we retrieve the results of [5]. We also note that the sign of the (conditional) covariance between two consecutive price moves does not depend on the initial configuration of the bid and ask queues and is given by the sign of $p_{cont} + p'_{cont} - 1$. We also note that the quantities p_{cont}, p'_{cont} can be computed up to the knowledge of the quantities $p_1^{up}(n_b, n_a)$ which computation was discussed in the previous section. The quantities $F(\pm\delta, t)$ can be computed up to the knowledge of the law of τ , which is known up to the use of a numerical procedure to invert the characteristic functions of σ_a and σ_b , together with the results of Proposition .4.

Proof. The results follow from elementary calculations in a similar way to what is done in [5]. Indeed, we have:

$$\begin{pmatrix} p_n^{up}(b, a) & 1 - p_n^{up}(b, a) \end{pmatrix} = \begin{pmatrix} p_1^{up}(b, a) & 1 - p_1^{up}(b, a) \end{pmatrix} \begin{pmatrix} p_{cont} & 1 - p_{cont} \\ 1 - p'_{cont} & p'_{cont} \end{pmatrix}^{n-1}$$

We also have:

$$\begin{pmatrix} p_{cont} & 1 - p_{cont} \\ 1 - p'_{cont} & p'_{cont} \end{pmatrix} = S \begin{pmatrix} 1 & 0 \\ 0 & p_{cont} + p'_{cont} - 1 \end{pmatrix} S^{-1}$$

with:

$$S = \begin{pmatrix} 1 & -\frac{1-p_{cont}}{1-p'_{cont}} \\ 1 & 1 \end{pmatrix}$$

□

4 Diffusion Limit of the Price Process

In [5] it is assumed that $f(i, j) = \tilde{f}(i, j) = f(j, i)$ in order to make the price increments X_n independent and identically distributed. In fact, this assumption can be entirely relaxed. Indeed, as we mentioned above, $(X_n, \tau_n)_{n \geq 0}$ is in fact a Markov renewal process and therefore we can use the related theory to compute the diffusion limit of the price process. The results of this section generalize the results of Section 4 in [5].

4.1 Balanced Order Flow case: $P^a(1, 1) = P^a(-1, -1)$ and $P^b(1, 1) = P^b(-1, -1)$

Throughout this section we make the assumption:

(A4) Using the notations of Proposition .1, the following holds:

$$\sum_{n=1}^{\infty} \sum_{p=1}^{\infty} \alpha^b(n) \alpha^a(p) f(n, p) < \infty, \quad \sum_{n=1}^{\infty} \sum_{p=1}^{\infty} \alpha^b(n) \alpha^a(p) \tilde{f}(n, p) < \infty.$$

Using Proposition .3, we obtain the following result generalizing lemma 1 in [5]:

Lemma .7. *Under assumption (A4), the following weak convergence holds as $n \rightarrow \infty$:*

$$\frac{1}{n \log(n)} \sum_{k=1}^n \tau_k \Rightarrow \tau^* := \sum_{n=1}^{\infty} \sum_{p=1}^{\infty} \alpha^b(n) \alpha^a(p) f^*(n, p),$$

$$\text{where } f^*(n, p) := \pi^* f(n, p) + (1 - \pi^*) \tilde{f}(n, p).$$

Proof. We have:

$$\frac{1}{n \log(n)} \sum_{k=1}^n \tau_k = \sum_{i \in \{-\delta, \delta\}} \frac{N_i(n)}{n} \frac{\log(N_i(n))}{\log(n)} \frac{1}{N_i(n) \log(N_i(n))} \sum_{k=1}^{N_i(n)} \tau_{p(k, i)},$$

where for $i \in \{-\delta, \delta\}$, $N_i(n)$ represents the number of times that $X_{k-1} = i$ for $1 \leq k \leq n$; and $\{p(k, i) : k \geq 1\}$ the successive indexes for which $X_{k-1} = i$. By the standard theory of Markov Chains, we have for $i \in \{-\delta, \delta\}$:

$$\frac{N_i(n)}{n} \xrightarrow{a.e.} \pi^*(i),$$

and therefore we have $\frac{\log(N_i(n))}{\log(n)} \xrightarrow{a.e.} 1$. We recall that $\pi^*(\delta) := \pi^*$, and $\pi^*(-\delta) = 1 - \pi^*$. For fixed $i \in \{-\delta, \delta\}$, the random variables $\{\tau_{p(k,i)} : k \geq 1\}$ are i.i.d. with distribution $F(i, \cdot)$, and with tail index equal to 1 (by Proposition .3). Using [5] (Lemma 1) together with Proposition .3, we get that:

$$\begin{aligned} \frac{1}{n \log(n)} \sum_{k=1}^n \tau_{p(k,\delta)} &\Rightarrow \sum_{n=1}^{\infty} \sum_{p=1}^{\infty} \alpha^b(n) \alpha^a(p) f(n, p), \\ \frac{1}{n \log(n)} \sum_{k=1}^n \tau_{p(k,-\delta)} &\Rightarrow \sum_{n=1}^{\infty} \sum_{p=1}^{\infty} \alpha^b(n) \alpha^a(p) \tilde{f}(n, p). \end{aligned}$$

The latter convergence holds in probability and we finally have:

$$\frac{1}{n \log(n)} \sum_{k=1}^n \tau_k \xrightarrow{P} \pi^* \sum_{n=1}^{\infty} \sum_{p=1}^{\infty} \alpha^b(n) \alpha^a(p) f(n, p) + (1 - \pi^*) \sum_{n=1}^{\infty} \sum_{p=1}^{\infty} \alpha^b(n) \alpha^a(p) \tilde{f}(n, p).$$

□

Let $s^* := \delta(2\pi^* - 1)$. Using the previous lemma .7, we obtain the following diffusion limit for the renormalized price process $s_{tn \log(n)}$:

Proposition .8. *Under assumption (A4), the renormalized price process $s_{tn \log(n)}$ satisfies the following weak convergence in the Skorokhod topology ([14]):*

$$\begin{aligned} \left(\frac{s_{tn \log(n)}}{n}, t \geq 0 \right) &\xrightarrow{n \rightarrow \infty} \left(\frac{s^* t}{\tau^*}, t \geq 0 \right), \\ \left(\frac{s_{tn \log(n)} - N_{tn \log(n)} s^*}{\sqrt{n}}, t \geq 0 \right) &\xrightarrow{n \rightarrow \infty} \frac{\sigma}{\sqrt{\tau^*}} W, \end{aligned}$$

where W is a standard Brownian motion and σ is given by:

$$\sigma^2 = 4\delta^2 \left(\frac{1 - p'_{cont} + \pi^*(p'_{cont} - p_{cont})}{(p_{cont} + p'_{cont} - 2)^2} - \pi^*(1 - \pi^*) \right).$$

Remark .9. *If $p'_{cont} = p_{cont} = \pi^* = \frac{1}{2}$ as in [5], we find $s^* = 0$ and $\sigma = \delta$ as in [5]. If $p'_{cont} = p_{cont} = p$, we have $\pi^* = \frac{1}{2}$, $s^* = 0$ and:*

$$\sigma^2 = \delta^2 \frac{p}{1-p}.$$

Proof. Because $m(\pm\delta) := \mathbb{E}[\tau_n | X_{n-1} = \pm\delta] = +\infty$ by Proposition .3, we cannot directly apply the well-known invariance principle results for semi-Markov processes. Denote for $t \in \mathbb{R}_+$:

$$R_n := \sum_{k=1}^n (X_k - s^*), \quad U_n(t) := n^{-1/2} [(1 - \lambda_{n,t})R_{\lfloor nt \rfloor} + \lambda_{n,t}R_{\lfloor nt \rfloor + 1}],$$

where $\lambda_{n,t} := nt - \lfloor nt \rfloor$. We can show, following a martingale method similar to [16] (section 3), that we have the following weak convergence in the Skorokhod topology:

$$(U_n(t), t \geq 0) \xrightarrow{n \rightarrow \infty} \sigma W,$$

where W is a standard Brownian motion, and σ is given by:

$$\sigma^2 = \sum_{i \in \{-\delta, \delta\}} \pi^*(i)v(i),$$

where for $i \in \{-\delta, \delta\}$:

$$\begin{aligned} v(i) &:= b(i)^2 + p(i)(g(-i) - g(i))^2 - 2b(i)p(i)(g(-i) - g(i)), \\ b(i) &:= i - s^*, \\ p(\delta) &:= 1 - p_{cont}, \quad p(-\delta) := 1 - p'_{cont}, \end{aligned}$$

and (the vector) g is given by:

$$g = (P + \Pi^* - I)^{-1}b,$$

where Π^* is the matrix with rows equal to $(\pi^* \quad 1 - \pi^*)$. After completing the calculations we get:

$$\sigma^2 = 4\delta^2 \left(\frac{1 - p'_{cont} + \pi^*(p'_{cont} - p_{cont})}{(p_{cont} + p'_{cont} - 2)^2} - \pi^*(1 - \pi^*) \right).$$

For the sake of exhaustivity we also give the explicit expression for g :

$$\begin{aligned} g(\delta) &= \delta \frac{p'_{cont} - p_{cont} + 2(1 - \pi^*)}{p_{cont} + p'_{cont} - 2} - s^*, \\ g(-\delta) &= \delta \frac{p'_{cont} - p_{cont} - 2\pi^*}{p_{cont} + p'_{cont} - 2} - s^*. \end{aligned}$$

Indeed, to show the above convergence of U_n , we observe that we can write R_n as the sum of a \mathcal{F}_n -martingale M_n and a bounded process:

$$R_n = M_n + \underbrace{g(X_n) - g(X_0) + X_n - X_0}_{unif.bounded}, \quad M_n := \sum_{k=1}^n b(X_{k-1}) - g(X_k) + g(X_{k-1}),$$

where $\mathcal{F}_n := \sigma(\tau_k, X_k : k \leq n)$ and $X_0 := 0$. The process M_n is a martingale because g is the unique solution of the following Poisson equation, since $\Pi^*b = 0$:

$$[P - I]g = b.$$

The rest of the proof for the convergence of U_n follows exactly [16] (section 3).

We proved earlier (lemma .7) that:

$$\frac{T_n}{n \log(n)} \Rightarrow \tau^*,$$

where $T_n := \sum_{k=1}^n \tau_k$. Since the Markov renewal process $(X_n, \tau_n)_{n \geq 0}$ is regular (because the state space is finite), we get $N_t \rightarrow \infty$ a.s. and therefore:

$$\frac{T_{N_t}}{N_t \log(N_t)} \Rightarrow \tau^*.$$

Observing that $T_{N_t} \leq t \leq T_{N_t+1}$ a.s., we get:

$$\frac{T_{N_t}}{N_t \log N_t} \leq \frac{t}{N_t \log N_t} \leq \frac{(N_t + 1) \log(N_t + 1)}{N_t \log N_t} \frac{T_{N_t+1}}{(N_t + 1) \log(N_t + 1)},$$

and therefore:

$$\frac{t}{N_t \log(N_t)} \Rightarrow \tau^*.$$

Let $t_n := tn \log(n)$. We would like to show as in [5], equation (17) that:

$$N_{t_n} \stackrel{P}{\sim} \frac{nt}{\tau^*}.$$

We have denoted by $A_n \stackrel{P}{\sim} B_n$ iff $P - \lim \frac{A_n}{B_n} = 1$. We denote as in [5] $\rho : (1, \infty) \rightarrow (1, \infty)$ to be the inverse function of $t \log(t)$, and we note that $\rho(t) \stackrel{t \rightarrow \infty}{\sim} \frac{t}{\log(t)}$. The first equivalence in [5], equation (17): $N_{t_n} \stackrel{P}{\sim} \rho\left(\frac{t_n}{\tau^*}\right)$ is not obvious. Indeed, we have $N_{t_n} \log(N_{t_n}) \stackrel{P}{\sim} \frac{t_n}{\tau^*}$, and we would like to conclude that $N_{t_n} = \rho(N_{t_n} \log(N_{t_n})) \stackrel{P}{\sim} \rho\left(\frac{t_n}{\tau^*}\right)$. The latter implication is not true for every function ρ , in particular if ρ was exponential. Nevertheless, in our case, it is true because $\rho(t) \stackrel{t \rightarrow \infty}{\sim} \frac{t}{\log(t)}$, and therefore for any functions f, g going to $+\infty$ as $t \rightarrow \infty$:

$$\frac{\rho(f(t))}{\rho(g(t))} \stackrel{t \rightarrow \infty}{\sim} \frac{f(t) \log(g(t))}{g(t) \log(f(t))}.$$

Therefore we see that if $f(t) \stackrel{t \rightarrow \infty}{\sim} g(t)$, then by property of the logarithm $\log(f(t)) \stackrel{t \rightarrow \infty}{\sim} \log(g(t))$ and therefore $\rho(f(t)) \stackrel{t \rightarrow \infty}{\sim} \rho(g(t))$. This allows us to conclude as in [5] that:

$$\frac{N_{t_n}}{n} \stackrel{P}{\sim} \frac{t}{\tau^*}.$$

Therefore, we can make a change of time as in [16], Corollary 3.19 (see also [1], section 14), and denoting $\alpha_n(t) := \frac{N_{tn}}{n}$, we obtain the following weak convergence in the Skorohod topology:

$$(U_n(\alpha_n(t)), t \geq 0) \Rightarrow (\sigma W_{\frac{t}{\tau^*}}, t \geq 0),$$

that is to say:

$$\left(\frac{stn \log(n) - N_{tn} \log(n) s^*}{\sqrt{n}}, t \geq 0 \right) \Rightarrow \frac{\sigma}{\sqrt{\tau^*}} W.$$

The law of large numbers result comes from the fact that $\frac{N_{tn}}{n} \xrightarrow{P} \frac{t}{\tau^*}$, together with the following fact (strong law of large numbers for Markov chains):

$$\frac{1}{n} \sum_{k=1}^n X_k \rightarrow s^* \quad a.e.$$

□

4.2 Other cases: either $P^a(1, 1) < P^a(-1, -1)$ or $P^b(1, 1) < P^b(-1, -1)$

In this case, we know by Proposition .3 that the conditional expectations $\mathbb{E}[\tau_k | q_0^b = n_b, q_0^a = n_a]$ are finite. Denoting the conditional expectations $m(\pm\delta) := \mathbb{E}[\tau_k | X_{k-1} = \pm\delta]$, we have:

$$\begin{aligned} m(\delta) &= \sum_{p=1}^{\infty} \sum_{n=1}^{\infty} \mathbb{E}[\tau_k | q_0^b = n, q_0^a = p] f(n, p), \\ m(-\delta) &= \sum_{p=1}^{\infty} \sum_{n=1}^{\infty} \mathbb{E}[\tau_k | q_0^b = n, q_0^a = p] \tilde{f}(n, p). \end{aligned}$$

Throughout this section we will need the following assumption:

(A5) Using the previous notations, the following holds:

$$m(\pm\delta) < \infty.$$

For example, the above assumption is satisfied if the support of the distributions f and \tilde{f} is compact, which is the case in practice. We obtain the following diffusion limit result as a classical consequence of invariance principle results for semi-Markov processes (see e.g. [16], section 3):

Proposition .10. *Under assumption (A5), the renormalized price process s_{nt} satisfies the following convergence in the Skorokhod topology:*

$$\begin{aligned} \left(\frac{s_{nt}}{n}, t \geq 0 \right) &\xrightarrow{n \rightarrow \infty} \left(\frac{s^* t}{m_\tau}, t \geq 0 \right) \quad a.e., \\ \left(\frac{s_{nt} - N_{nt} s^*}{\sqrt{n}}, t \geq 0 \right) &\xrightarrow{n \rightarrow \infty} \frac{\sigma}{\sqrt{m_\tau}} W, \end{aligned}$$

where W is a standard Brownian motion, σ is given in Proposition .8 and:

$$m_\tau := \sum_{i \in \{-\delta, \delta\}} \pi^*(i) m(i) = \pi^* m(\delta) + (1 - \pi^*) m(-\delta).$$

Proof. This is an immediate consequence of strong law of large numbers and invariance principle results for Markov renewal processes satisfying $m(\pm\delta) < \infty$ (see e.g. [16] section 3). In the previous article [16], the proof of the invariance principle is carried on using a martingale method similar to the one of the proof of proposition .8. \square

5 Numerical Results

In this section, we present calibration results which illustrate and justify our approach.

In [5], it is assumed that the queue changes V_k^b, V_k^a do not depend on their previous values V_{k-1}^b, V_{k-1}^a . Empirically, it is found that $\mathbb{P}[V_k^b = 1] \approx \mathbb{P}[V_k^b = -1] \approx 1/2$ (and similarly for the ask side). Here, we challenge this assumption by estimating and comparing the probabilities $P(-1, 1)$ Vs. $P(1, 1)$ on the one side and $P(-1, -1)$ Vs. $P(1, -1)$ on the other side to check whether or not they are approximately equal to each other, for both the ask and the bid. We also give - for both the bid and ask - the estimated probabilities $\mathbb{P}[V_k = 1]$, $\mathbb{P}[V_k = -1]$ that we call respectively $P(1)$, $P(-1)$, to check whether or not they are approximately equal to 1/2 as in [5].

The results below correspond to the 5 stocks Amazon, Apple, Google, Intel, Microsoft on June 21st 2012¹. The probabilities are estimated using the strong law of large numbers. We also give for indicative purposes the average time between order arrivals (in milliseconds (ms)) as well as the average number of stocks per order.

| | Amazon | | Apple | | Google | | Intel | | Microsoft | |
|-----------------------------|--------|-----|-------|-----|--------|------|-------|-----|-----------|-----|
| | Bid | Ask | Bid | Ask | Bid | Ask | Bid | Ask | Bid | Ask |
| Avg time btw. orders (ms) | 910 | 873 | 464 | 425 | 1123 | 1126 | 116 | 133 | 130 | 113 |
| Avg nb. of stocks per order | 100 | 82 | 90 | 82 | 84 | 71 | 502 | 463 | 587 | 565 |

Average time between orders (ms) & Average number of stocks per order. June 21st 2012.

| | Amazon | | Apple | | Google | | Intel | | Microsoft | |
|-------------|--------|-------|-------|------|--------|------|-------|-------|-----------|-------|
| | Bid | Ask | Bid | Ask | Bid | Ask | Bid | Ask | Bid | Ask |
| $P(1, 1)$ | 0.48 | 0.57 | 0.50 | 0.55 | 0.48 | 0.53 | 0.55 | 0.61 | 0.63 | 0.60 |
| $P(-1, 1)$ | 0.46 | 0.42 | 0.40 | 0.42 | 0.46 | 0.49 | 0.44 | 0.40 | 0.36 | 0.41 |
| $P(-1, -1)$ | 0.54 | 0.58 | 0.60 | 0.58 | 0.54 | 0.51 | 0.56 | 0.60 | 0.64 | 0.59 |
| $P(1, -1)$ | 0.52 | 0.43 | 0.50 | 0.45 | 0.52 | 0.47 | 0.45 | 0.39 | 0.37 | 0.40 |
| $P(1)$ | 0.47 | 0.497 | 0.44 | 0.48 | 0.47 | 0.51 | 0.495 | 0.505 | 0.49 | 0.508 |
| $P(-1)$ | 0.53 | 0.503 | 0.56 | 0.52 | 0.53 | 0.49 | 0.505 | 0.495 | 0.51 | 0.492 |

Estimated transition probabilities of the Markov Chains V_k^b, V_k^a . June 21st 2012.

¹The data was taken from the webpage <https://lobster.wiwi.hu-berlin.de/info/DataSamples.php>

Findings: First of all, we find as in [5] that for all stocks, $\mathbb{P}[V_k = 1] \approx \mathbb{P}[V_k = -1] \approx 1/2$, except maybe in the case of Apple Bid. It is worth mentioning that we always have $P(1) < P(-1)$ except in 3 cases: Google Ask, Intel Ask and Microsoft Ask. Nevertheless, in these cases, $P(1)$ and $P(-1)$ are very close to each other and so they could be considered to fall into the case $P(1) = P(-1)$ of [5]. These 3 cases also correspond to the only 3 cases where $P(1, 1) > P(-1, -1)$, which is contrary to our assumption $P(1, 1) \leq P(-1, -1)$. Nevertheless, in these 3 cases, $P(1, 1)$ and $P(-1, -1)$ are very close to each other so we can consider them to fall into the case $P(1, 1) = P(-1, -1)$.

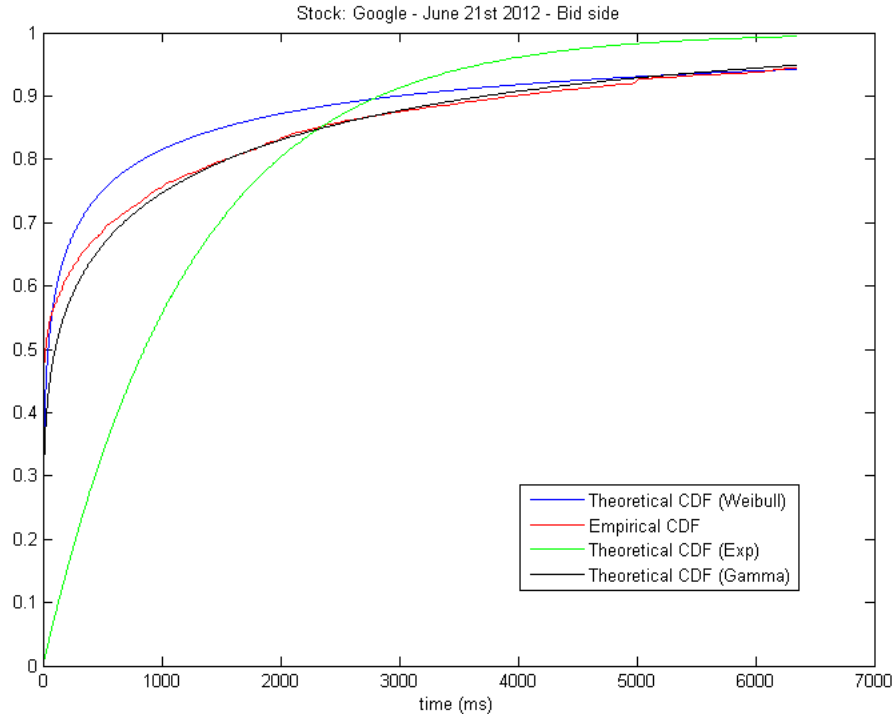
More importantly, we notice that the probabilities $P(-1, 1)$, $P(1, 1)$ can be significantly different from each other - and similarly for the probabilities $P(-1, -1)$, $P(1, -1)$ - which justifies the use of a Markov Chain structure for the random variables $\{V_k^b\}, \{V_k^a\}$. This phenomenon is particularly visible for example on Microsoft (Bid+Ask), Intel (Bid+Ask), Apple (Bid+Ask) or Amazon Ask. Further, regarding the comparison of $P(1, 1)$ and $P(-1, -1)$, it turns out that they are often very similar, except in the cases Amazon Bid, Apple Bid, Google Bid.

The second assumption of [5] that we would like to challenge is the assumed exponential distribution of the order arrival times T_k^a, T_k^b . To this end, on the same data set as used to estimate the transition probabilities $P^a(i, j)$, $P^b(i, j)$, we calibrate the empirical c.d.f.'s $H^a(i, j, \cdot)$, $H^b(i, j, \cdot)$ to the Gamma and Weibull distributions (which are generalizations of the exponential distribution). We recall that the p.d.f.'s of these distributions are given by:

$$f_{Gamma}(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}} 1_{x>0},$$

$$f_{Weibull}(x) = \frac{k}{\theta} \left(\frac{x}{\theta}\right)^{k-1} e^{-\left(\frac{x}{\theta}\right)^k} 1_{x>0}.$$

Here, $k > 0$ and $\theta > 0$ represent respectively the shape and the scale parameter. The variable k is dimensionless, whereas θ will be expressed in ms^{-1} . We perform a maximum likelihood estimation of the Weibull and Gamma parameters for each one of the empirical distributions $H^a(i, j, \cdot)$, $H^b(i, j, \cdot)$ (together with a 95 % confidence interval for the parameters). As we can see on the tables below, the shape parameter k is always significantly different than 1 (~ 0.1 to 0.3), which indicates that the exponential distribution is not rich enough to fit our observations. To illustrate this, we present below the empirical c.d.f. of $H(1, -1)$ in the case of Google Bid, and we see that Gamma and Weibull allow to fit the empirical c.d.f. in a much better way than Exponential.



$H(1, -1)$ - Google Bid - June 21st 2012.

We summarize our calibration results in the tables below.

| Amazon Bid | $H(1, 1)$ | $H(1, -1)$ | $H(-1, -1)$ | $H(-1, 1)$ |
|-------------------|------------------------|------------------------|------------------------|------------------------|
| Weibull θ | 99.1 (90.2-109.0) | 185.5 (171.3-200.8) | 87.7 (80.1-96.0) | 87.0 (78.7-96.1) |
| Weibull k | 0.279 (0.274-0.285) | 0.323 (0.317-0.329) | 0.285 (0.280-0.290) | 0.258 (0.253-0.263) |
| Gamma θ | 4927 (4618-5257) | 4321 (4075-4582) | 4712 (4423-5019) | 5965 (5589-6366) |
| Gamma k | 0.179 (0.174-0.184) | 0.215 (0.209-0.220) | 0.179 (0.175-0.184) | 0.165 (0.161-0.169) |

Amazon Bid: Fitted Weibull and Gamma parameters. 95 % confidence intervals in brackets. June 21st 2012.

| Amazon Ask | $H(1, 1)$ | $H(1, -1)$ | $H(-1, -1)$ | $H(-1, 1)$ |
|-------------------|------------------------|------------------------|------------------------|------------------------|
| Weibull θ | 80.8 (74.4-87.7) | 197.8 (181.9-215.1) | 57.9 (52.8-63.4) | 137.0 (124.2-151.2) |
| Weibull k | 0.274 (0.269-0.278) | 0.324 (0.317-0.330) | 0.279 (0.274-0.285) | 0.276 (0.270-0.281) |
| Gamma θ | 4732 (4475-5004) | 4623 (4345-4919) | 3845 (3609-4095) | 5879 (5502-6283) |
| Gamma k | 0.174 (0.170-0.178) | 0.215 (0.209-0.221) | 0.173 (0.168-0.177) | 0.181 (0.176-0.186) |

Amazon Ask: Fitted Weibull and Gamma parameters. 95 % confidence intervals in brackets. June 21st 2012.

| Apple Bid | $H(1, 1)$ | $H(1, -1)$ | $H(-1, -1)$ | $H(-1, 1)$ |
|------------------|------------------------|------------------------|------------------------|------------------------|
| Weibull θ | 75.9 (71.6-80.5) | 180.9 (172.6-189.7) | 31.5 (29.5-33.6) | 78.2 (73.4-83.3) |
| Weibull k | 0.317 (0.313-0.321) | 0.400 (0.394-0.405) | 0.271 (0.267-0.274) | 0.300 (0.296-0.304) |
| Gamma θ | 2187 (2094-2284) | 1860 (1787-1935) | 2254 (2157-2355) | 2711 (2592-2835) |
| Gamma k | 0.206 (0.202-0.210) | 0.276 (0.271-0.282) | 0.168 (0.165-0.171) | 0.196 (0.192-0.199) |

Apple Bid: Fitted Weibull and Gamma parameters. 95 % confidence intervals in brackets. June 21st 2012.

| Apple Ask | $H(1, 1)$ | $H(1, -1)$ | $H(-1, -1)$ | $H(-1, 1)$ |
|------------------|------------------------|------------------------|------------------------|------------------------|
| Weibull θ | 46.6 (44.1-49.2) | 152.5 (145.5-159.8) | 27.7 (26.0-29.6) | 95.5 (90.0-101.5) |
| Weibull k | 0.298 (0.294-0.301) | 0.394 (0.388-0.399) | 0.271 (0.267-0.275) | 0.308 (0.303-0.312) |
| Gamma θ | 2019 (1942-2099) | 1666 (1603-1732) | 1995 (1907-2087) | 2740 (2624-2861) |
| Gamma k | 0.189 (0.186-0.192) | 0.271 (0.266-0.277) | 0.168 (0.165-0.171) | 0.204 (0.200-0.208) |

Apple Ask: Fitted Weibull and Gamma parameters. 95 % confidence intervals in brackets. June 21st 2012.

| Google Bid | $H(1, 1)$ | $H(1, -1)$ | $H(-1, -1)$ | $H(-1, 1)$ |
|-------------------|------------------------|------------------------|------------------------|------------------------|
| Weibull θ | 113.9 (102.8-126.2) | 158.5 (143.4-175.3) | 67.9 (60.6-76.0) | 56.8 (50.5-63.8) |
| Weibull k | 0.276 (0.270-0.282) | 0.284 (0.278-0.290) | 0.261 (0.255-0.266) | 0.246 (0.241-0.251) |
| Gamma θ | 6720 (6263-7210) | 6647 (6204-7122) | 6381 (5913-6886) | 7025 (6517-7571) |
| Gamma k | 0.174 (0.169-0.179) | 0.185 (0.180-0.191) | 0.160 (0.155-0.165) | 0.151 (0.147-0.156) |

Google Bid: Fitted Weibull and Gamma parameters. 95 % confidence intervals in brackets. June 21st 2012.

| Google Ask | $H(1, 1)$ | $H(1, -1)$ | $H(-1, -1)$ | $H(-1, 1)$ |
|-------------------|------------------------|------------------------|------------------------|------------------------|
| Weibull θ | 196.7 (180.6-214.2) | 271.6 (248.5-296.8) | 38.1 (33.8-43.0) | 57.0 (51.3-63.3) |
| Weibull k | 0.290 (0.285-0.295) | 0.310 (0.303-0.316) | 0.258 (0.253-0.264) | 0.263 (0.258-0.268) |
| Gamma θ | 6081 (5734-6450) | 6571 (6165-7003) | 4304 (3971-4664) | 4698 (4380-5040) |
| Gamma k | 0.195 (0.190-0.200) | 0.209 (0.203-0.215) | 0.156 (0.151-0.161) | 0.164 (0.159-0.168) |

Google Ask: Fitted Weibull and Gamma parameters. 95 % confidence intervals in brackets. June 21st 2012.

| Intel Bid | $H(1, 1)$ | $H(1, -1)$ | $H(-1, -1)$ | $H(-1, 1)$ |
|------------------|------------------------|------------------------|------------------------|------------------------|
| Weibull θ | 2.76 (2.66-2.86) | 2.56 (2.45-2.67) | 3.33 (3.21-3.45) | 2.01 (1.92-2.10) |
| Weibull k | 0.227 (0.226-0.229) | 0.226 (0.225-0.228) | 0.267 (0.265-0.269) | 0.209 (0.208-0.211) |
| Gamma θ | 1016 (991-1040) | 912 (888-937) | 543 (530-557) | 1093 (1063-1124) |
| Gamma k | 0.129 (0.128-0.130) | 0.130 (0.129-0.131) | 0.151 (0.150-0.152) | 0.120 (0.119-0.121) |

Intel Bid: Fitted Weibull and Gamma parameters. 95 % confidence intervals in brackets. June 21st 2012.

| Intel Ask | $H(1, 1)$ | $H(1, -1)$ | $H(-1, -1)$ | $H(-1, 1)$ |
|------------------|------------------------|------------------------|------------------------|------------------------|
| Weibull θ | 1.33 (1.28-1.38) | 5.46 (5.21-5.73) | 4.63 (4.45-4.80) | 5.15 (4.90-5.41) |
| Weibull k | 0.235 (0.234-0.237) | 0.231 (0.230-0.233) | 0.256 (0.254-0.257) | 0.225 (0.224-0.227) |
| Gamma θ | 705 (688-723) | 1219 (1183-1256) | 884 (862-907) | 1305 (1266-1345) |
| Gamma k | 0.126 (0.125-0.127) | 0.137 (0.136-0.139) | 0.146 (0.144-0.147) | 0.133 (0.132-0.135) |

Intel Ask: Fitted Weibull and Gamma parameters. 95 % confidence intervals in brackets. June 21st 2012.

| Microsoft Bid | $H(1, 1)$ | $H(1, -1)$ | $H(-1, -1)$ | $H(-1, 1)$ |
|----------------------|------------------------|------------------------|------------------------|------------------------|
| Weibull θ | 0.79 (0.76-0.82) | 2.98 (2.83-3.13) | 2.68 (2.59-2.78) | 2.64 (2.50-2.78) |
| Weibull k | 0.215 (0.214-0.217) | 0.221 (0.219-0.223) | 0.259 (0.257-0.260) | 0.211 (0.209-0.213) |
| Gamma θ | 1012 (987-1039) | 1315 (1274-1358) | 664 (648-681) | 1488 (1440-1537) |
| Gamma k | 0.112 (0.111-0.113) | 0.125 (0.124-0.127) | 0.142 (0.141-0.143) | 0.120 (0.118-0.121) |

Microsoft Bid: Fitted Weibull and Gamma parameters. 95 % confidence intervals in brackets. June 21st 2012.

| Microsoft Ask | $H(1, 1)$ | $H(1, -1)$ | $H(-1, -1)$ | $H(-1, 1)$ |
|----------------------|------------------------|------------------------|------------------------|------------------------|
| Weibull θ | 0.85 (0.82-0.89) | 1.57 (1.50-1.64) | 2.07 (2.00-2.15) | 1.43 (1.36-1.50) |
| Weibull k | 0.218 (0.217-0.219) | 0.223 (0.222-0.225) | 0.259 (0.258-0.261) | 0.210 (0.208-0.211) |
| Gamma θ | 1004 (980-1028) | 1081 (1051-1112) | 574 (560-588) | 1138 (1105-1171) |
| Gamma k | 0.113 (0.112-0.114) | 0.121 (0.120-0.122) | 0.140 (0.139-0.141) | 0.116 (0.115-0.117) |

Microsoft Ask: Fitted Weibull and Gamma parameters. 95 % confidence intervals in brackets. June 21st 2012.

6 Conclusion and Future Work

In this paper, we introduced a semi-Markovian modeling of limit order books in order to match empirical observations. We extended the model of [5] in the following ways:

- i) inter-arrival times between book events (limit orders, market orders, order cancellations) are allowed to have an arbitrary distribution.
- ii) the arrival of a new book event at the bid or the ask and its corresponding inter-arrival time are allowed to depend on the nature of the previous event.

In order to do so, both the bid and ask queues are driven by Markov renewal processes. It results from these chosen dynamics that the price process can be expressed as a functional of another Markov renewal process, which we characterized explicitly. In this context, we obtained probabilistic results such as the duration until the next price change, the probability of price increase and the characterization of the Markov renewal process driving the stock price process (section 3). In section 4, we obtained diffusion limit results for the stock price process generalizing those of [5]. Finally, we presented in section 5 calibration results on real market data in order to illustrate and justify our approach.

References

- [1] Billingsley, P. *Convergence of probability measures*, John Wiley & Sons, Inc., 1999.
- [2] A. Cartea, and S. Jaimungal, *Optimal Execution with Limit and Market Orders*, Quantitative Finance, forthcoming.
- [3] K. J. Cohen, R. M. Conroy and S. F. Maier, *Order flow and the quality of the market*, in Market Making and the Changing Structure of the Securities Industry (1985).
- [4] R. Cont, S. Stoikov and R. Talreja, *A stochastic model for order book dynamics*, Operations Research, 58 (2010), pp. 549-563.
- [5] R. Cont and A. de Larrard, *Price dynamics in a Markovian limit order book market*, SIAM Journal for Financial Mathematics, 4 (2013), No 1, pp. 1-25.

- [6] I. Domowitz and J. Wang, *Auctions as algorithms*, J. Economic Dynamics and Control, 18 (1994), pp. 29-60.
- [7] P. Fodra and H. Pham, *Semi Markov model for market microstructure*, ArXiv:1305v1 [q-fin.TR], 1 May (2013).
- [8] P. Fodra and H. Pham, *High frequency trading and asymptotics for small risk aversion in a Markov renewal model*, arXiv:1310.1765v2 [q-fin.TR], 4 Jan (2015).
- [9] M. Garman, *Market microstructure*, J. Financial Economics, 3 (1976), pp. 257-275.
- [10] N. Limnios and G. Oprisan, *Semi-Markov Processes and Reliability*, Birkhauser, 2001.
- [11] H. Luckock, *A steady-state model of a continuous double auction*, Quant. Finance, 3 (2003), pp. 385-404.
- [12] H. Mendelson, *Market behavior in a clearing house*, Econometrica, 50 (1982), pp. 1505-1524.
- [13] S. Predoiu, G. Shaikhet and S. Shreve, *Optimal execution in a general one-sided limit order book*, SIAM J. Financial Math., 2(1) (2011), pp. 183-212.
- [14] A. Skorokhod, *Studies in the Theory of Random Processes*, Addison-Wesley, Reading, Mass., 1965 (Reprinted by Dover Publications, NY).
- [15] E. Smith, J.D. Farmer, L. Gillemot and S. Krishnamurthy, *Statistical theory of the continuous double auction*, Quant. Finance, 3 (2003), pp. 481-514.
- [16] N. Vadori and A. Swishchuk, *Strong Law of Large Numbers and Central Limit Theorems for Functionals of Inhomogeneous Semi-Markov Processes*, Stochastic Analysis and Applications, 33:2 (2015), pp. 213-243.